Estimation of Toeplitz Covariance Matrices in Large Dimensional Regime with Application to Source Detection

Julia Vinogradova, Romain Couillet, and Walid Hachem

Abstract

In this article, we derive concentration inequalities for the spectral norm of two classical sample estimators of large dimensional Toeplitz covariance matrices, demonstrating in particular their asymptotic almost sure consistence. The consistency is then extended to the case where the aggregated matrix of time samples is corrupted by a rank one (or more generally, low rank) matrix. As an application of the latter, the problem of source detection in the context of large dimensional sensor networks within a temporally correlated noise environment is studied. As opposed to standard procedures, this application is performed online, *i.e.* without the need to possess a learning set of pure noise samples.

Index Terms

Covariance matrix, concentration inequalities, correlated noise, source detection.

I. INTRODUCTION

Let $(v_t)_{t \in \mathbb{Z}}$ be a complex circularly symmetric Gaussian stationary process with zero mean and covariance function $(r_k)_{k \in \mathbb{Z}}$ with $r_k = \mathbb{E}[v_{t+k} v_t^*]$ and $r_k \to 0$ as $k \to \infty$. We observe N independent copies of $(v_t)_{t \in \mathbb{Z}}$ over the time window $t \in \{0, \ldots, T-1\}$, and stack the observations in a matrix $V_T = [v_{n,t}]_{n,t=0}^{N-1,T-1}$. This matrix can be written as $V_T = W_T R_T^{1/2}$, where $W_T \in \mathbb{C}^{N \times T}$ has independent $\mathcal{CN}(0,1)$ (standard circularly symmetric complex Gaussian) entries and $R_T^{1/2}$ is any square root of the Hermitian nonnegative definite Toeplitz $T \times T$ matrix

$$R_T \triangleq [r_{i-j}]_{0 \le i,j \le T-1} = \begin{bmatrix} r_0 & r_1 & \ldots & r_{T-1} \\ r_{-1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & r_1 \\ r_{1-T} & \ldots & r_{-1} & r_0 \end{bmatrix}.$$

A classical problem in signal processing is to estimate R_T from the observation of V_T. With the growing importance of multi-antenna array processing, there has recently been a renewed interest for this estimation problem in the regime of large system dimensions, *i.e.* for both N and T large.

At the core of the various estimation methods for R_T are the biased and unbiased estimates $\hat{r}^b_{k,T}$ and $\hat{r}^u_{k,T}$ for r_k, respectively, defined by

$$\hat{r}^b_{k,T} = \frac{1}{NT} \sum_{n=0}^{N-1} \sum_{t=0}^{T-1} v_{n,t+k} v^*_{n,t} \mathbb{1}_{0 \leq t+k \leq T-1}$$

$$\hat{r}^u_{k,T} = \frac{1}{N(T-|k|)} \sum_{n=0}^{N-1} \sum_{t=0}^{T-1} v_{n,t+k} v^*_{n,t} \mathbb{1}_{0 \leq t+k \leq T-1}$$

where $\mathbb{1}_A$ is the indicator function on the set A. Depending on the relative rate of growth of N and T, the matrices $\widehat{R}^b_T = [\hat{r}^b_{i-j,T}]_{0 \leq i,j \leq T-1}$ and $\widehat{R}^u_T = [\hat{r}^u_{i-j,T}]_{0 \leq i,j \leq T-1}$ may not satisfy $\|R_T - \widehat{R}^b_T\| \xrightarrow{\text{a.s.}} 0$ or $\|R_T - \widehat{R}^u_T\| \xrightarrow{\text{a.s.}} 0$. An important drawback of the biased entry-wise estimate lies in its inducing a general asymptotic bias in $\widehat{R}^b_T$; as for the unbiased entry-wise estimate, it may induce too much inaccuracy in the top-right and bottom-left entries of $\widehat{R}^u_T$. The estimation paradigm followed in the recent literature generally consists instead in building banded or tapered versions of $\widehat{R}^b_T$ or $\widehat{R}^u_T$ (*i.e.* by weighting down or discarding a certain number of entries away from the diagonal), exploiting there the rate of decrease of r_k as $k \to \infty$ [1], [2], [3], [4], [5], [6]. Such estimates use the fact that $\|R_T - R_{\gamma(T),T}\| \to 0$ with $R_{\gamma,T} = [[R_T]_{i,j} \mathbb{1}_{|i-j| \leq \gamma}]$ for some well-chosen functions $\gamma(T)$ (usually satisfying $\gamma(T) \to \infty$ and $\gamma(T)/T \to 0$) and restrict the study to the consistent estimation of $R_{\gamma(T),T}$. The aforementioned articles concentrate in particular on choices of functions $\gamma(T)$ that ensure optimal rates of convergence of $\|R_T - \widehat{R}_{\gamma(T),T}\|$ for the banded or tapered estimate $\widehat{R}_{\gamma(T),T}$. These procedures, although theoretically optimal, however suffer from several practical limitations. First, they assume the *a priori* knowledge of the rate of decrease of r_k (and restrict these rates to specific classes). Then, even if this were indeed known in practice, being asymptotic in nature, the results do not provide explicit rules for selecting $\gamma(T)$ for practical finite values of N and T. Finally, the operations of banding and tapering do not guarantee the positive definiteness of the resulting covariance estimate.

In the present article, we consider instead that the only constraint about r_k is $\sum_{k=-\infty}^{\infty} |r_k| < \infty$ and estimate R_T from the standard (non-banded and non-tapered) estimates $\widehat{R}^b_T$ and $\widehat{R}^u_T$. The consistence of these estimates, in general invalid, shall be enforced here by the choice $N, T \to \infty$ with $N/T \to c \in (0, \infty)$. This setting is more practical in applications as long as both the finite values N and T are sufficiently large and of similar order of magnitude. Another context where a non banded Toeplitz rectification of the estimated covariance matrix leads to a consistent estimate in the spectral norm is studied in [7].

Our specific contribution lies in the establishment of concentration inequalities for the random variables $\|R_T - \widehat{R}^b_T\|$ and $\|R_T - \widehat{R}^u_T\|$. It is shown specifically that, for all $x > 0$, $-\log \mathbb{P}[\|R_T - \widehat{R}^b_T\| > x] = O(T)$ and $-\log \mathbb{P}[\|R_T - \widehat{R}^u_T\| > x] = O(T/\log T)$. Aside from the consistence in norm, this implies as a corollary that, as long as $\limsup_T \|R_T^{-1}\| < \infty$, for T large enough, $\widehat{R}^u_T$ is positive definite with outstanding probability ($\widehat{R}^b_T$ is nonnegative definite by construction).

For application purposes, the results are then extended to the case where V_T is changed into $V_T + P_T$ for a rank-one matrix P_T. Under some conditions on the right-eigenspaces of P_T, we show that the concentration inequalities hold identically. The application is that of a single source detection (modeled through P_T) by an array of N sensors embedded in a temporally correlated noise (modeled by V_T). To proceed to detection, R_T is estimated from $V_T + P_T$ as $\widehat{R}^b_T$ or $\widehat{R}^u_T$, which is used as a whitening matrix, before applying a generalized

likelihood ratio test (GLRT) procedure on the whitened observation. Simulations corroborate the theoretical consistence of the test.

The remainder of the article is organized as follows. The concentration inequalities for both biased and unbiased estimates are exposed in Section II. The generalization to the rank-one perturbation model is presented in Section III and applied in the practical context of source detection in Section IV.

Notations: The superscript $(\cdot)^{\mathsf{H}}$ denotes Hermitian transpose, $\|X\|$ stands for the spectral norm for a matrix and Euclidean norm for a vector, and $\|\cdot\|_\infty$ is the sup norm of a function. The notations $\mathcal{N}(a,\sigma^2)$ and $\mathcal{CN}(a,\sigma^2)$ represent the real and complex circular Gaussian distributions with mean a and variance σ^2. For $x \in \mathbb{C}^m$, $D_x = \mathrm{diag}(x) = \mathrm{diag}(x_0,\ldots,x_{m-1})$ is the diagonal matrix having on its diagonal the elements of the vector x. For $x = [x_{-(m-1)},\ldots,x_{m-1}]^{\mathsf{T}} \in \mathbb{C}^{2m+1}$, the matrix $\mathcal{T}(x) \in \mathbb{C}^{m \times m}$ is the Toeplitz matrix built from x with entries $[\mathcal{T}(x)]_{i,j} = x_{j-i}$. The notations $\Re(\cdot)$ and $\Im(\cdot)$ stand for the real and the imaginary parts respectively.

II. Performance of the covariance matrix estimators

A. Model, assumptions, and results

Let $(r_k)_{k\in\mathbb{Z}}$ be a doubly infinite sequence of covariance coefficients. For any $T \in \mathbb{N}$, let $R_T = \mathcal{T}(r_{-(T-1)},\ldots,r_{T-1})$, a Hermitian nonnegative definite matrix. Given $N = N(T) > 0$, consider the matrix model

$$V_T = [v_{n,t}]_{n,t=0}^{N-1,T-1} = W_T R_T^{1/2} \tag{1}$$

where $W_T = [w_{n,t}]_{n,t=0}^{N-1,T-1}$ has independent $\mathcal{CN}(0,1)$ entries. It is clear that $r_k = \mathbb{E}[v_{n,t+k}v_{n,t}^*]$ for any t, k, and $n \in \{0,\ldots,N-1\}$.

In the following, we shall make the two assumptions below.

Assumption 1. *The covariance coefficients r_k are absolutely summable and $r_0 \neq 0$.*

With this assumption, the covariance function

$$\boldsymbol{\Upsilon}(\lambda) \triangleq \sum_{k=-\infty}^{\infty} r_k e^{-\imath k\lambda}, \quad \lambda \in [0, 2\pi)$$

is continuous on the interval $[0, 2\pi]$. Since $\|R_T\| \leq \|\boldsymbol{\Upsilon}\|_\infty$ (see *e.g.* [8, Lemma 4.1]), Assumption 1 implies that $\sup_T \|R_T\| < \infty$.

We assume the following asymptotic regime which will be simply denoted as "$T \to \infty$":

Assumption 2. *$T \to \infty$ and $N/T \to c > 0$.*

Our objective is to study the performance of two estimators of the covariance function frequently considered in the literature. These estimators are defined as

$$\hat{r}_{k,T}^b = \frac{1}{NT} \sum_{n=0}^{N-1}\sum_{t=0}^{T-1} v_{n,t+k}v_{n,t}^* \mathbb{1}_{0 \leq t+k \leq T-1} \tag{2}$$

$$\hat{r}_{k,T}^u = \frac{1}{N(T-|k|)} \sum_{n=0}^{N-1}\sum_{t=0}^{T-1} v_{n,t+k}v_{n,t}^* \mathbb{1}_{0 \leq t+k \leq T-1}. \tag{3}$$

Since $\mathbb{E}\hat{r}^b_{k,T} = (1 - |k|/T)r_k$ and $\mathbb{E}\hat{r}^u_{k,T} = r_k$, the estimate $\hat{r}^b_{k,T}$ is biased while $\hat{r}^u_{k,T}$ is unbiased. Let also

$$\widehat{R}^b_T \triangleq \mathcal{T}\left(\hat{r}^b_{-(T-1),T}, \ldots, \hat{r}^b_{(T-1),T}\right) \tag{4}$$

$$\widehat{R}^u_T \triangleq \mathcal{T}\left(\hat{r}^u_{-(T-1),T}, \ldots, \hat{r}^u_{(T-1),T}\right). \tag{5}$$

A well known advantage of $\widehat{R}^b_T$ over $\widehat{R}^u_T$ as an estimate of R_T is its structural nonnegative definiteness. In this section, results on the spectral behavior of these matrices are provided under the form of concentration inequalities on $\|\widehat{R}^b_T - R_T\|$ and $\|\widehat{R}^u_T - R_T\|$:

Theorem 1. *Let Assumptions 1 and 2 hold true and let $\widehat{R}^b_T$ be defined as in* (4). *Then, for any $x > 0$,*

$$\mathbb{P}\left[\left\|\widehat{R}^b_T - R_T\right\| > x\right] \leq \exp\left(-cT\left(\frac{x}{\|\mathbf{\Upsilon}\|_\infty} - \log\left(1 + \frac{x}{\|\mathbf{\Upsilon}\|_\infty}\right) + o(1)\right)\right)$$

where $o(1)$ is with respect to T and depends on x.

Theorem 2. *Let Assumptions 1 and 2 hold true and let $\widehat{R}^u_T$ be defined as in* (5). *Then, for any $x > 0$,*

$$\mathbb{P}\left[\left\|\widehat{R}^u_T - R_T\right\| > x\right] \leq \exp\left(-\frac{cTx^2}{4\|\mathbf{\Upsilon}\|^2_\infty \log T}(1 + o(1))\right)$$

where $o(1)$ is with respect to T and depends on x.

A consequence of these theorems, obtained by the Borel-Cantelli lemma, is that $\|\widehat{R}^b_T - R_T\| \to 0$ and $\|\widehat{R}^u_T - R_T\| \to 0$ almost surely as $T \to \infty$.

The slower rate of decrease of $T/\log(T)$ in the unbiased estimator exponent may be interpreted by the increased inaccuracy in the estimates of r_k for values of k close to $T - 1$.

We now turn to the proofs of Theorems 1 and 2, starting with some basic mathematical results that will be needed throughout the proofs.

B. Some basic mathematical facts

Lemma 1. *For $x, y \in \mathbb{C}^m$ and $A \in \mathbb{C}^{m \times m}$,*

$$\left|x^\mathsf{H}Ax - y^\mathsf{H}Ay\right| \leq \|A\|\left(\|x\| + \|y\|\right)\|x - y\|.$$

Proof:

$$\left|x^\mathsf{H}Ax - y^\mathsf{H}Ay\right| = \left|x^\mathsf{H}Ax - y^\mathsf{H}Ax + y^\mathsf{H}Ax - y^\mathsf{H}Ay\right|$$

$$\leq \left|(x - y)^\mathsf{H}Ax\right| + \left|y^\mathsf{H}A(x - y)\right|$$

$$\leq \|A\|\left(\|x\| + \|y\|\right)\|x - y\|.$$

∎

Lemma 2. *Let $X_0, \ldots, X_{M-1}$ be independent $\mathcal{CN}(0, 1)$ random variables. Then, for any $x > 0$,*

$$\mathbb{P}\left[\frac{1}{M}\sum_{m=0}^{M-1}\left(|X_m|^2 - 1\right) > x\right] \leq \exp\left(-M(x - \log(1 + x))\right).$$

Proof: This is a classical Chernoff bound. Indeed, given $\xi \in (0,1)$, we have by the Markov inequality

$$\mathbb{P}\Big[M^{-1}\sum_{m=0}^{M-1}(|X_m|^2-1) > x\Big] = \mathbb{P}\left[\exp\left(\xi\sum_{m=0}^{M-1}|X_m|^2\right) > \exp\xi M(x+1)\right]$$

$$\leq \exp(-\xi M(x+1))\mathbb{E}\left[\exp\left(\xi\sum_{m=0}^{M-1}|X_m|^2\right)\right]$$

$$= \exp\left(-M\left(\xi(x+1)+\log(1-\xi)\right)\right)$$

since $\mathbb{E}\left[\exp(\xi|X_m|^2)\right] = 1/(1-\xi)$. The result follows upon minimizing this expression with respect to ξ. ∎

C. Biased estimator: proof of Theorem 1

Define

$$\widehat{\Upsilon}_T^b(\lambda) \triangleq \sum_{k=-(T-1)}^{T-1} \hat{r}_{k,T}^b e^{\imath k\lambda}$$

$$\Upsilon_T(\lambda) \triangleq \sum_{k=-(T-1)}^{T-1} r_k e^{\imath k\lambda}.$$

Since $\widehat{R}_T^b - R_T$ is a Toeplitz matrix, from [8, Lemma 4.1],

$$\left\|\widehat{R}_T^b - R_T\right\| \leq \sup_{\lambda\in[0,2\pi)}\left|\widehat{\Upsilon}_T^b(\lambda) - \Upsilon_T(\lambda)\right| \leq \sup_{\lambda\in[0,2\pi)}\left|\widehat{\Upsilon}_T^b(\lambda) - \mathbb{E}\widehat{\Upsilon}_T^b(\lambda)\right| + \sup_{\lambda\in[0,2\pi)}\left|\mathbb{E}\widehat{\Upsilon}_T^b(\lambda) - \Upsilon_T(\lambda)\right|.$$

By Kronecker's lemma ([9, Lemma 3.21]), the rightmost term at the right-hand side satisfies

$$\left|\mathbb{E}\widehat{\Upsilon}_T^b(\lambda) - \Upsilon_T(\lambda)\right| \leq \sum_{k=-(T-1)}^{T-1} \frac{|kr_k|}{T} \xrightarrow[T\to\infty]{} 0. \tag{6}$$

In order to deal with the term $\sup_{\lambda\in[0,2\pi)}|\widehat{\Upsilon}_T^b(\lambda) - \mathbb{E}\widehat{\Upsilon}_T^b(\lambda)|$, two ingredients will be used. The first one is the following lemma (proven in Appendix A1):

Lemma 3. *The following facts hold:*

$$\widehat{\Upsilon}_T^b(\lambda) = d_T(\lambda)^\mathsf{H}\frac{V_T^\mathsf{H}V_T}{N}d_T(\lambda)$$

$$\mathbb{E}\widehat{\Upsilon}_T^b(\lambda) = d_T(\lambda)^\mathsf{H}R_T d_T(\lambda)$$

where $d_T(\lambda) = 1/\sqrt{T}\left[1, e^{-\imath\lambda}, \ldots, e^{-\imath(T-1)\lambda}\right]^\mathsf{T}$.

The second ingredient is a Lipschitz property of the function $\|d_T(\lambda) - d_T(\lambda')\|$ seen as a function of λ. From the inequality $|e^{-\imath t\lambda} - e^{-\imath t\lambda'}| \leq t|\lambda - \lambda'|$, we indeed have

$$\|d_T(\lambda) - d_T(\lambda')\| = \sqrt{\frac{1}{T}\sum_{t=0}^{T-1}|e^{-\imath t\lambda} - e^{-\imath t\lambda'}|^2} \leq \frac{T|\lambda-\lambda'|}{\sqrt{3}}. \tag{7}$$

Now, denoting by $\lfloor\cdot\rfloor$ the floor function and choosing $\beta > 2$, define $\mathcal{I} = \left\{0, \ldots, \lfloor T^\beta\rfloor - 1\right\}$. Let $\lambda_i =$

$2\pi\frac{i}{\lfloor T^\beta\rfloor}$, $i\in\mathcal{I}$, be a regular discretization of the interval $[0,2\pi]$. We write

$$\sup_{\lambda\in[0,2\pi)}\left|\widehat{\Upsilon}_T^b(\lambda)-\mathbb{E}\widehat{\Upsilon}_T^b(\lambda)\right|$$

$$\leq\max_{i\in\mathcal{I}}\sup_{\lambda\in[\lambda_i,\lambda_{i+1}]}\left(\left|\widehat{\Upsilon}_T^b(\lambda)-\widehat{\Upsilon}_T^b(\lambda_i)\right|+\left|\widehat{\Upsilon}_T^b(\lambda_i)-\mathbb{E}\widehat{\Upsilon}_T^b(\lambda_i)\right|+\left|\mathbb{E}\widehat{\Upsilon}_T^b(\lambda_i)-\mathbb{E}\widehat{\Upsilon}_T^b(\lambda)\right|\right)$$

$$\leq\max_{i\in\mathcal{I}}\sup_{\lambda\in[\lambda_i,\lambda_{i+1}]}\left|\widehat{\Upsilon}_T^b(\lambda)-\widehat{\Upsilon}_T^b(\lambda_i)\right|+\max_{i\in\mathcal{I}}\left|\widehat{\Upsilon}_T^b(\lambda_i)-\mathbb{E}\widehat{\Upsilon}_T^b(\lambda_i)\right|+\max_{i\in\mathcal{I}}\sup_{\lambda\in[\lambda_i,\lambda_{i+1}]}\left|\mathbb{E}\widehat{\Upsilon}_T^b(\lambda_i)-\mathbb{E}\widehat{\Upsilon}_T^b(\lambda)\right|$$

$$\triangleq\chi_1+\chi_2+\chi_3.$$

With the help of Lemma 3 and (7), we shall provide concentration inequalities on the random terms χ_1 and χ_2 and a bound on the deterministic term χ_3. This is the purpose of the three following lemmas. Herein and in the remainder, C denotes a positive constant independent of T. This constant can change from an expression to another.

Lemma 4. *There exists a constant $C>0$ such that for any $x>0$ and any T large enough,*

$$\mathbb{P}\left[\chi_1>x\right]\leq\exp\left(-cT^2\left(\frac{xT^{\beta-2}}{C\|\mathbf{\Upsilon}\|_\infty}-\log\frac{xT^{\beta-2}}{C\|\mathbf{\Upsilon}\|_\infty}-1\right)\right).$$

Proof: Using Lemmas 3 and 1 along with (7), we have

$$\left|\widehat{\Upsilon}_T^b(\lambda)-\widehat{\Upsilon}_T^b(\lambda_i)\right|=\left|d_T(\lambda)^{\mathsf{H}}\frac{V_T^{\mathsf{H}}V_T}{N}d_T(\lambda)-d_T(\lambda_i)^{\mathsf{H}}\frac{V_T^{\mathsf{H}}V_T}{N}d_T(\lambda_i)\right|$$

$$\leq 2N^{-1}\|d_T(\lambda)-d_T(\lambda_i)\|\,\|R_T\|\,\|W_T^{\mathsf{H}}W_T\|$$

$$\leq C|\lambda-\lambda_i|\|\mathbf{\Upsilon}\|_\infty\,\|W_T^{\mathsf{H}}W_T\|.$$

From $\|W_T^{\mathsf{H}}W_T\|\leq\mathrm{Tr}(W_T^{\mathsf{H}}W_T)$ and Lemma 2, assuming T large enough so that $f(x,T)\triangleq xT^{\beta-1}/(CN\|\mathbf{\Upsilon}\|_\infty)$ satisfies $f(x,T)\geq 1$, we then obtain

$$\mathbb{P}\left[\chi_1>x\right]\leq\mathbb{P}\left[C\|\mathbf{\Upsilon}\|_\infty T^{-\beta}\sum_{t=0}^{T-1}\sum_{n=0}^{N-1}|w_{n,t}|^2>x\right]$$

$$=\mathbb{P}\left[\frac{1}{NT}\sum_{n,t}(|w_{n,t}|^2-1)>f(x,T)-1\right]$$

$$\leq\exp(-NT(f(x,T)-\log f(x,T)-1)).$$

■

Lemma 5. *The following inequality holds*

$$\mathbb{P}\left[\chi_2>x\right]\leq 2T^\beta\exp\left(-cT\left(\frac{x}{\|\mathbf{\Upsilon}\|_\infty}-\log\left(1+\frac{x}{\|\mathbf{\Upsilon}\|_\infty}\right)\right)\right).$$

Proof: From the union bound we obtain:

$$\mathbb{P}\left[\chi_2>x\right]\leq\sum_{i=0}^{\lfloor T^\beta\rfloor-1}\mathbb{P}\left[\left|\widehat{\Upsilon}_T^b(\lambda_i)-\mathbb{E}\widehat{\Upsilon}_T^b(\lambda_i)\right|>x\right].$$

We shall bound each term of the sum separately. Since

$$\mathbb{P}\left[\left|\widehat{\Upsilon}_T^b(\lambda_i) - \mathbb{E}\widehat{\Upsilon}_T^b(\lambda_i)\right| > x\right] = \mathbb{P}\left[\widehat{\Upsilon}_T^b(\lambda_i) - \mathbb{E}\widehat{\Upsilon}_T^b(\lambda_i) > x\right] + \mathbb{P}\left[-\left(\widehat{\Upsilon}_T^b(\lambda_i) - \mathbb{E}\widehat{\Upsilon}_T^b(\lambda_i)\right) > x\right]$$

it will be enough to deal with the first right-hand side term as the second one is treated similarly. Let $\eta_T(\lambda_i) \triangleq W_T q_T(\lambda_i) = [\eta_{0,T}(\lambda_i), \ldots, \eta_{N-1,T}(\lambda_i)]^\mathsf{T}$ where $q_T(\lambda_i) \triangleq R_T^{1/2} d_T(\lambda_i)$. Observe that $\eta_{k,T}(\lambda_i) \sim \mathcal{CN}(0, \|q_T(\lambda_i)\|^2 I_N)$. We know from Lemma 3 that

$$\widehat{\Upsilon}_T^b(\lambda_i) - \mathbb{E}\widehat{\Upsilon}_T^b(\lambda_i) = \frac{1}{N}\left(\|\eta_T(\lambda_i)\|^2 - \mathbb{E}\|\eta_T(\lambda_i)\|^2\right). \tag{8}$$

From (8) and Lemma 2, we therefore get

$$\mathbb{P}\left[\widehat{\Upsilon}_T^b(\lambda_i) - \mathbb{E}\widehat{\Upsilon}_T^b(\lambda_i) > x\right] \le \exp\left(-N\left(\frac{x}{\|q_T(\lambda_i)\|^2} - \log\left(1 + \frac{x}{\|q_T(\lambda_i)\|^2}\right)\right)\right).$$

Noticing that $\|q_T(\lambda_i)\|^2 \le \|\mathbf{\Upsilon}\|_\infty$ and that the function $f(x) = x - \log\left(1 + x\right)$ is increasing for $x > 0$, we get the result. ∎

Finally, the bound for the deterministic term χ_3 is provided by the following lemma:

Lemma 6. $\chi_3 \le C\|\mathbf{\Upsilon}\|_\infty T^{-\beta+1}$.

Proof: From Lemmas 3 and 1 along with (7), we obtain

$$\left|\mathbb{E}\widehat{\Upsilon}_T^b(\lambda) - \mathbb{E}\widehat{\Upsilon}_T^b(\lambda_i)\right| = \left|d_T(\lambda)^\mathsf{H} R_T d_T(\lambda) - d_T(\lambda_i)^\mathsf{H} R_T d_T(\lambda_i)\right|$$

$$\le 2\|R_T\|\|d_T(\lambda) - d_T(\lambda_i)\|$$

$$\le C\|\mathbf{\Upsilon}\|_\infty |\lambda - \lambda_i| T.$$

From $\max_{i\in\mathcal{I}} \sup_{\lambda\in[\lambda_i,\lambda_{i+1}]} |\lambda - \lambda_i| = \lambda_{i+1} - \lambda_i = T^{-\beta}$ we get the result. ∎

We now complete the proof of Theorem 1. From (6) and Lemma 6, we get

$$\mathbb{P}\left[\left\|\widehat{R}_T^b - R_T\right\| > x\right] = \mathbb{P}\left[\chi_1 + \chi_2 > x + o(1)\right].$$

Given a parameter $\epsilon_T \in [0, 1]$, we can write (with some slight notation abuse)

$$\mathbb{P}\left[\chi_1 + \chi_2 > x + o(1)\right] \le \mathbb{P}\left[\chi_1 > x\epsilon_T\right] + \mathbb{P}\left[\chi_2 > x(1 - \epsilon_T) + o(1)\right].$$

With the results of Lemmas 4 and 5, setting $\epsilon_T = 1/T$, we get

$$\mathbb{P}\left[\chi_1 + \chi_2 > x + o(1)\right] \le \mathbb{P}\left[\chi_1 > \frac{x}{T}\right] + \mathbb{P}\left[\chi_2 > x(1 - \frac{x}{T}) + o(1)\right]$$

$$\le \exp\left(-cT^2\left(\frac{xT^{\beta-3}}{C\|\mathbf{\Upsilon}\|_\infty} - \log\frac{xT^{\beta-3}}{C\|\mathbf{\Upsilon}\|_\infty} - 1\right)\right)$$

$$+ \exp\left(-cT\left(\frac{x\left(1 - \frac{1}{T}\right)}{\|\mathbf{\Upsilon}\|_\infty} - \log\left(1 + \frac{x\left(1 - \frac{1}{T}\right)}{\|\mathbf{\Upsilon}\|_\infty}\right) + o(1)\right)\right)$$

$$= \exp\left(-cT\left(\frac{x}{\|\mathbf{\Upsilon}\|_\infty} - \log\left(1 + \frac{x}{\|\mathbf{\Upsilon}\|_\infty}\right) + o(1)\right)\right)$$

since $\beta > 2$.

D. Unbiased estimator: proof of Theorem 2

The proof follows basically the same main steps as for Theorem 1 with an additional difficulty due to the scaling terms $1/(T - |k|)$.

Defining the function

$$\widehat{\Upsilon}_T^u(\lambda) \triangleq \sum_{k=-(T-1)}^{T-1} \hat{r}_{k,T}^u e^{ik\lambda}$$

we have

$$\left\| \widehat{R}_T^u - R_T \right\| \leq \sup_{\lambda \in [0,2\pi)} \left| \widehat{\Upsilon}_T^u(\lambda) - \Upsilon_T(\lambda) \right| = \sup_{\lambda \in [0,2\pi)} \left| \widehat{\Upsilon}_T^u(\lambda) - \mathbb{E}\widehat{\Upsilon}_T^u(\lambda) \right|$$

since $\Upsilon_T(\lambda) = \mathbb{E}\widehat{\Upsilon}_T^u(\lambda)$, the estimates $\hat{r}_{k,T}^u$ being unbiased.

In order to deal with the right-hand side of this expression, we need the following analogue of Lemma 3, borrowed from [7] and proven here in Appendix B1.

Lemma 7. *The following fact holds:*

$$\widehat{\Upsilon}_T^u(\lambda) = d_T(\lambda)^{\mathsf{H}} \left(\frac{V_T^{\mathsf{H}} V_T}{N} \odot B_T \right) d_T(\lambda)$$

where $\odot$ is the Hadamard product of matrices and where

$$B_T \triangleq \left[\frac{T}{T - |i - j|} \right]_{0 \leq i,j \leq T-1}.$$

In order to make $\widehat{\Upsilon}_T^u(\lambda)$ more tractable, we rely on the following lemma which can be proven by direct calculation.

Lemma 8. *Let $x, y \in \mathbb{C}^m$ and $A, B \in C^{m \times m}$. Then*

$$x^{\mathsf{H}}(A \odot B)y = \text{Tr}(D_x^{\mathsf{H}} A D_y B^{\mathsf{T}})$$

where we recall $D_x = \text{diag}(x)$ and $D_y = \text{diag}(y)$.

Denoting

$$D_T(\lambda) \triangleq \text{diag}(d_T(\lambda)) = \frac{1}{\sqrt{T}} \text{diag}(1, e^{i\lambda}, \ldots, e^{i(T-1)\lambda})$$

$$Q_T(\lambda) \triangleq R_T^{1/2} D_T(\lambda) B_T D_T(\lambda)^{\mathsf{H}} (R_T^{1/2})^{\mathsf{H}}$$

we get from Lemmas 7 and 8

$$\begin{aligned}
\widehat{\Upsilon}_T^u(\lambda) &= \frac{1}{N} \text{Tr}(D_T(\lambda)^{\mathsf{H}} (R_T^{1/2})^{\mathsf{H}} W_T^{\mathsf{H}} W_T R_T^{1/2} D_T(\lambda) B_T) \\
&= \frac{1}{N} \text{Tr}(W_T Q_T(\lambda) W_T^{\mathsf{H}}) \\
&= \frac{1}{N} \sum_{n=0}^{N-1} w_n^{\mathsf{H}} Q_T(\lambda) w_n
\end{aligned} \tag{9}$$

where w_i^{H} is such that $W_T = [w_0^{\mathsf{H}}, \ldots, w_{N-1}^{\mathsf{H}}]$.

Compared to the biased case, the main difficulty lies here in the fact that the matrices B_T/T and $Q_T(\lambda)$

have unbounded spectral norm as $T \to \infty$. The following lemma, proven in Appendix B2, provides some information on the spectral behavior of these matrices that will be used subsequently.

Lemma 9. *The matrix B_T satisfies*

$$\|B_T\| \leq \sqrt{2}T(\sqrt{\log T} + C). \tag{10}$$

For any $\lambda \in [0, 2\pi)$, the eigenvalues $\sigma_0, \ldots, \sigma_{T-1}$ of the matrix $Q(\lambda)$ satisfy the following inequalities:

$$\sum_{t=0}^{T-1} \sigma_t^2 \leq 2\|\mathbf{\Upsilon}\|_\infty^2 \log T + C \tag{11}$$

$$\max_t |\sigma_t| \leq \sqrt{2}\|\mathbf{\Upsilon}\|_\infty (\log T)^{1/2} + C \tag{12}$$

$$\sum_{t=0}^{T-1} |\sigma_t|^3 \leq C((\log T)^{3/2} + 1) \tag{13}$$

where the constant C is independent of λ.

We shall also need the following easily shown Lipschitz property of the function $\|D_T(\lambda) - D_T(\lambda')\|$:

$$\|D_T(\lambda) - D_T(\lambda')\| \leq \sqrt{T}|\lambda - \lambda'|. \tag{14}$$

We now enter the core of the proof of Theorem 2. Choosing $\beta > 2$, let $\lambda_i = 2\pi \frac{i}{\lfloor T^\beta \rfloor}$, $i \in \mathcal{I}$, be a regular discretization of the interval $[0, 2\pi]$ with $\mathcal{I} = \{0, \ldots, \lfloor T^\beta \rfloor - 1\}$. We write

$$\sup_{\lambda \in [0,2\pi)} \left| \widehat{\Upsilon}_T^u(\lambda) - \mathbb{E}\widehat{\Upsilon}_T^u(\lambda) \right| \leq \max_{i \in \mathcal{I}} \sup_{\lambda \in [\lambda_i, \lambda_{i+1}]} \left| \widehat{\Upsilon}_T^u(\lambda) - \widehat{\Upsilon}_T^u(\lambda_i) \right| + \max_{i \in \mathcal{I}} \left| \widehat{\Upsilon}_T^u(\lambda_i) - \mathbb{E}\widehat{\Upsilon}_T^u(\lambda_i) \right|$$

$$+ \max_{i \in \mathcal{I}} \sup_{\lambda \in [\lambda_i, \lambda_{i+1}]} \left| \mathbb{E}\widehat{\Upsilon}_T^u(\lambda_i) - \mathbb{E}\widehat{\Upsilon}_T^u(\lambda) \right|$$

$$\triangleq \chi_1 + \chi_2 + \chi_3.$$

Our task is now to provide concentration inequalities on the random terms χ_1 and χ_2 and a bound on the deterministic term χ_3.

Lemma 10. *There exists a constant $C > 0$ such that, if T is large enough, the following inequality holds:*

$$\mathbb{P}[\chi_1 > x] \leq \exp\left(-cT^2 \left(\frac{xT^{\beta-2}}{C\sqrt{\log T}} - \log \frac{xT^{\beta-2}}{C\sqrt{\log T}} - 1\right)\right).$$

Proof: From Equation (9), we have

$$\left| \widehat{\Upsilon}_T^u(\lambda) - \widehat{\Upsilon}_T^u(\lambda_i) \right| = \frac{1}{N} \left| \sum_{n=0}^{N-1} w_n^{\mathsf{H}} \left(Q_T(\lambda) - Q_T(\lambda_i)\right) w_n \right|$$

$$\leq \frac{1}{N} \sum_{n=0}^{N-1} \left| w_n^{\mathsf{H}} \left(Q_T(\lambda) - Q_T(\lambda_i)\right) w_n \right|$$

$$\leq \frac{1}{N} \|Q_T(\lambda) - Q_T(\lambda_i)\| \sum_{n=0}^{N-1} \|w_n\|^2.$$

The norm above further develops as

$$\|Q_T(\lambda) - Q_T(\lambda_i)\|$$

$$\leq \|R_T\| \|D_T(\lambda)B_T D_T(\lambda)^{\mathsf{H}} - D_T(\lambda_i)B_T D_T(\lambda)^{\mathsf{H}} + D_T(\lambda_i)B_T D_T(\lambda)^{\mathsf{H}} - D_T(\lambda_i)B_T D_T(\lambda_i)^{\mathsf{H}}\|$$

$$\leq 2 \|D_T(\lambda)\| \|R_T\| \|B_T\| \|D_T(\lambda) - D_T(\lambda_i)\| \leq CT(\sqrt{\log T} + 1) |\lambda - \lambda_i|$$

where we used (10), (14), and $\|D_T(\lambda)\| = 1/\sqrt{T}$. Up to a change in C, we can finally write $\|Q_T(\lambda) - Q_T(\lambda_i)\| \leq CT^{1-\beta}\sqrt{\log T}$. Assume that $f(x,T) \triangleq xT^{\beta-2}/\left(C\sqrt{\log T}\right)$ satisfies $f(x,T) > 1$ (always possible for every fixed x by taking T large). Then we get by Lemma 2

$$\mathbb{P}\left[\chi_1 > x\right] \leq \mathbb{P}\left(CN^{-1}T^{1-\beta}\sqrt{\log T} \sum_{n,t} |w_{n,t}|^2 > x\right)$$

$$= \mathbb{P}\left(\frac{1}{NT} \sum_{n,t} (|w_{n,t}|^2 - 1) > f(x,T) - 1\right)$$

$$\leq \exp\left(-NT\left(f(x,T) - \log\left(f(x,T)\right) - 1\right)\right).$$

∎

The most technical part of the proof is to control the term χ_2, which we handle hereafter.

Lemma 11. *The following inequality holds:*

$$\mathbb{P}\left[\chi_2 > x\right] \leq \exp\left(-\frac{cx^2 T}{4 \|\mathbf{\Upsilon}\|_\infty^2 \log T}(1 + o(1))\right).$$

Proof: From the union bound we obtain:

$$\mathbb{P}\left[\chi_2 > x\right] \leq \sum_{i=0}^{\lfloor T^\beta \rfloor - 1} \mathbb{P}\left[\left|\widehat{\Upsilon}_T^u(\lambda_i) - \mathbb{E}\widehat{\Upsilon}_T^u(\lambda_i)\right| > x\right]. \tag{15}$$

Each term of the sum can be written

$$\mathbb{P}\left[\left|\widehat{\Upsilon}_T^u(\lambda_i) - \mathbb{E}\widehat{\Upsilon}_T^u(\lambda_i)\right| > x\right] = \mathbb{P}\left[\widehat{\Upsilon}_T^u(\lambda_i) - \mathbb{E}\widehat{\Upsilon}_T^u(\lambda_i) > x\right] + \mathbb{P}\left[-\left(\widehat{\Upsilon}_T^u(\lambda_i) - \mathbb{E}\widehat{\Upsilon}_T^u(\lambda_i)\right) > x\right].$$

We will deal with the term $\psi_i = \mathbb{P}\left[\widehat{\Upsilon}_T^u(\lambda_i) - \mathbb{E}\widehat{\Upsilon}_T^u(\lambda_i) > x\right]$, the term $\mathbb{P}\left[-\left(\widehat{\Upsilon}_T^u(\lambda_i) - \mathbb{E}\widehat{\Upsilon}_T^u(\lambda_i)\right) > x\right]$ being treated similarly. Let $Q_T(\lambda_i) = U_T \Sigma_T U_T^{\mathsf{H}}$ be a spectral factorization of the Hermitian matrix $Q_T(\lambda_i)$ with $\Sigma_T = \mathrm{diag}(\sigma_0, \ldots, \sigma_{T-1})$. Since U_T is unitary and W_T has independent $\mathcal{CN}(0,1)$ elements, we get from Equation (9)

$$\widehat{\Upsilon}_T^u(\lambda_i) \stackrel{\mathcal{L}}{=} \frac{1}{N} \sum_{n=0}^{N-1} w_n^{\mathsf{H}} \Sigma_T(\lambda_i) w_n = \frac{1}{N} \sum_{n=0}^{N-1} \sum_{t=0}^{T-1} |w_{n,t}|^2 \sigma_t \tag{16}$$

where $\stackrel{\mathcal{L}}{=}$ denotes equality in law. Since $\mathbb{E}\left[e^{a|X|^2}\right] = 1/(1-a)$ when $X \sim \mathcal{CN}(0,1)$ and $0 < a < 1$, we have

by Markov's inequality and from the independence of the variables $|w_{n,t}|^2$

$$
\begin{aligned}
\psi_i &= \mathbb{P}\left(\frac{1}{N} \sum_{n=0}^{N-1} \sum_{t=0}^{T-1} |w_{n,t}|^2 \sigma_t - \operatorname{Tr} Q_T(\lambda_i) > x \right) \\
&\leq \mathbb{E}\left[\exp\left(\frac{\tau}{N} \sum_{n,t} |w_{n,t}|^2 \sigma_t \right) \right] \exp\left(-\tau\left(x + \sum_{t=0}^{T-1} \sigma_t \right) \right) \\
&= \exp\left(-\tau\left(x + \sum_{t=0}^{T-1} \sigma_t \right) \right) \prod_{t=0}^{T-1} \left(1 - \frac{\sigma_t \tau}{N} \right)^{-N} \\
&= \exp\left(-\tau\left(x + \sum_{t=0}^{T-1} \sigma_t \right) - N \sum_{t=0}^{T-1} \log\left(1 - \frac{\sigma_t \tau}{N} \right) \right)
\end{aligned}
\tag{17}
$$

for any τ such that $0 \leq \tau < \min\limits_{0 \leq t \leq T-1} \frac{N}{\sigma_t}$. Writing $\log(1-x) = -x - \frac{x^2}{2} + R_3(x)$ with $|R_3(x)| \leq \frac{|x|^3}{3(1-\epsilon)^3}$ when $|x| < \epsilon < 1$, we get

$$
\begin{aligned}
\psi_i &\leq \exp\left(-\tau x + N \sum_{t=0}^{T-1} \left(\frac{\sigma_t^2 \tau^2}{2N^2} + R_3\left(\frac{\sigma_t \tau}{N} \right) \right) \right) \\
&\leq \exp\left(-N\left(\frac{\tau x}{N} - \frac{\tau^2}{2N^2} \sum_{t=0}^{T-1} \sigma_t^2 \right) \right) \exp\left(N \sum_{t=0}^{T-1} \left| R_3\left(\frac{\sigma_t \tau}{N} \right) \right| \right).
\end{aligned}
\tag{18}
$$

We shall manage this expression by using Lemma 9. In order to control the term $\exp(N \sum |R_3(\cdot)|)$, we make the choice

$$
\tau = \frac{axT}{\log T}
$$

where a is a parameter of order one to be optimized later. From (12) we get $\max_t \frac{\sigma_t \tau}{N} = O\left((\log T)^{-1/2} \right)$. Hence, for all T large, $\tau < \min_t \frac{N}{\sigma_t}$. Therefore, (17) is valid for this choice of τ and for T large. Moreover, for ϵ fixed and T large, $\frac{\sigma_t \tau}{N} < \epsilon < 1$ so that for these T

$$
N \sum_{t=0}^{T-1} \left| R_3\left(\frac{\sigma_t \tau}{N} \right) \right| \leq \frac{a^3 T^3 x^3}{3N^2 (1-\epsilon)^3 (\log T)^3} \sum_{t=0}^{T-1} |\sigma_t|^3 = O\left(T(\log T)^{-3/2} \right)
$$

from (13). Plugging the expression of τ in (18), we get

$$
\psi_i \leq \exp\left(-N\left(\frac{aTx^2}{(\log T)N} - \frac{a^2 T^2 x^2}{2N^2 (\log T)^2} \sum_{t=0}^{T-1} \sigma_t^2 \right) \right) \exp\left(C\left(T(\log T)^{-3/2} \right) \right).
$$

Using (11), we have

$$
\psi_i \leq \exp\left(-\frac{x^2 T}{\log T}\left(a - \frac{\|\Upsilon\|_\infty^2 a^2 T}{N} \right) \right) \exp\left(\frac{CT}{(\log T)^{3/2}} \right).
$$

The right hand side term is minimized for $a = \frac{N}{2T\|\Upsilon\|_\infty^2}$ which finally gives

$$
\psi_i \leq \exp\left(-\frac{Nx^2}{4\|\Upsilon\|_\infty^2 \log T}(1 + o(1)) \right).
$$

Combining the above inequality with (15) (which induces additional $o(1)$ terms in the argument of the exponential) concludes the lemma. ∎

Lemma 12. $\chi_3 \leq CT^{-\beta+2}\sqrt{\log T}$.

Proof: From Lemma 7, $\|R_T \odot B_T\| \le \|R_T\| \|B_T\|$ (see [10, Theorem 5.5.1]), and (7), we get:

$$\left| \mathbb{E}\widehat{\Upsilon}_T^u(\lambda_i) - \mathbb{E}\widehat{\Upsilon}_T^u(\lambda) \right| \le 2 \|d_T(\lambda) - d_T(\lambda_i)\| \|R_T\| \|B_T\| \le CT^2 |\lambda - \lambda_i| \|\mathbf{\Upsilon}\|_\infty \sqrt{\log T}.$$

■

Lemmas 10–12 show that $\mathbb{P}[\chi_2 > x]$ dominates the term $\mathbb{P}[\chi_1 > x]$ and that the term χ_3 is vanishing. Mimicking the end of the proof of Theorem 1, we obtain Theorem 2.

We conclude this section by an empirical evaluation by Monte Carlo simulations of $\mathbb{P}[\|\widehat{R}_T - R_T\| > x]$ (curves labeled Biased and Unbiased), with $\widehat{R}_T \in \{\widehat{R}_T^b, \widehat{R}_T^u\}$, $T = 2N$, $x = 2$. This is shown in Figure 1 against the theoretical exponential bounds of Theorems 1 and 2 (curves labeled Biased theory and Unbiased theory). We observe that the rates obtained in Theorems 1 and 2 are asymptotically close to optimal.
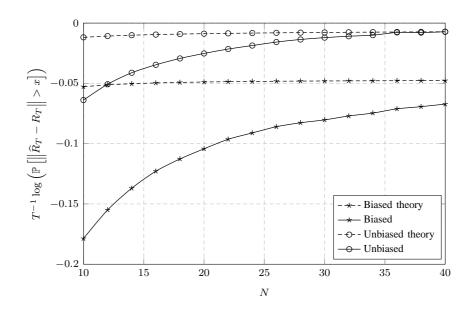

Figure 1. Error probability of the spectral norm for $x = 2$, $c = 0.5$, $[R_T]_{k,l} = a^{|k-l|}$ with $a = 0.6$.

III. COVARIANCE MATRIX ESTIMATORS FOR THE "SIGNAL PLUS NOISE" MODEL

A. Model, assumptions, and results

Consider now the following model:

$$Y_T = [y_{n,t}]_{\substack{0 \le n \le N-1 \\ 0 \le t \le T-1}} = P_T + V_T \tag{19}$$

where the $N \times T$ matrix V_T is defined in (1) and where P_T satisfies the following assumption:

Assumption 3. $P_T \triangleq \boldsymbol{h}_T \boldsymbol{s}_T^{\mathsf{H}} \Gamma_T^{1/2}$ *where* $\boldsymbol{h}_T \in \mathbb{C}^N$ *is a deterministic vector such that* $\sup_T \|\boldsymbol{h}_T\| < \infty$, *the vector* $\boldsymbol{s}_T = (s_0, \ldots, s_{T-1})^\mathsf{T} \in \mathbb{C}^T$ *is a random vector independent of* W_T *with the distribution* $\mathcal{CN}(0, I_T)$, *and* $\Gamma_T = [\gamma_{ij}]_{i,j=0}^{T-1}$ *is Hermitian nonnegative such that* $\sup_T \|\Gamma_T\| < \infty$.

We have here a model for a rank-one signal corrupted with a Gaussian spatially white and temporally correlated noise with stationary temporal correlations. Observe that the signal can also be temporally correlated.

Our purpose is still to estimate the noise correlation matrix R_T. To that end, we use one of the estimators (2) or (3) with the difference that the samples $v_{n,t}$ are simply replaced with the samples $y_{n,t}$. It turns out that these estimators are still consistent in spectral norm. Intuitively, P_T does not break the consistence of these estimators as it can be seen as a rank-one perturbation of the noise term V_T in which the subspace spanned by $(\Gamma^{1/2})^{\mathsf{H}} s_T$ is "delocalized" enough so as not to perturb much the estimators of R_T. In fact, we even have the following strong result.

Theorem 3. *Let Y_T be defined as in* (19) *and let Assumptions 1–3 hold. Define the estimates*

$$\hat{r}_{k,T}^{bp} = \frac{1}{NT} \sum_{n=0}^{N-1} \sum_{t=0}^{T-1} y_{n,t+k} y_{n,t}^* \mathbb{1}_{0 \le t+k \le T-1}$$

$$\hat{r}_{k,T}^{up} = \frac{1}{N(T-|k|)} \sum_{n=0}^{N-1} \sum_{t=0}^{T-1} y_{n,t+k} y_{n,t}^* \mathbb{1}_{0 \le t+k \le T-1}$$

and let

$$\widehat{R}_T^{bp} = \mathcal{T}(\hat{r}_{-(T-1),T}^{bp}, \cdots, \hat{r}_{(T-1),T}^{bp})$$

$$\widehat{R}_T^{up} = \mathcal{T}(\hat{r}_{-(T-1),T}^{up}, \cdots, \hat{r}_{(T-1),T}^{up}).$$

Then for any $x > 0$,

$$\mathbb{P}\left[\left\|\widehat{R}_T^{bp} - R_T\right\| > x\right] \le \exp\left(-cT\left(\frac{x}{\|\mathbf{\Upsilon}\|_\infty} - \log\left(1 + \frac{x}{\|\mathbf{\Upsilon}\|_\infty}\right) + o(1)\right)\right)$$

and

$$\mathbb{P}\left[\left\|\widehat{R}_T^{up} - R_T\right\| > x\right] \le \exp\left(-\frac{cTx^2}{4\|\mathbf{\Upsilon}\|_\infty^2 \log T}(1 + o(1))\right).$$

Before proving this theorem, some remarks are in order.

Remark 1. *Theorem 3 generalizes without difficulty to the case where P_T has a fixed rank $K > 1$. This captures the situation of $K \ll \min(N, T)$ sources.*

Remark 2. *Similar to the proofs of Theorems 1 and 2, the proof of Theorem 3 uses concentration inequalities for functionals of Gaussian random variables based on the moment generating function and the Chernoff bound. Exploiting instead McDiarmid's concentration inequality [11], it is possible to adapt Theorem 3 to s_T with bounded (instead of Gaussian) entries. This adaptation may account for discrete sources met in digital communication signals.*

B. Main elements of the proof of Theorem 3

We restrict the proof to the more technical part that concerns $\widehat{R}_T^{up}$. Defining

$$\widehat{\Upsilon}_T^{up}(\lambda) \triangleq \sum_{k=-(T-1)}^{T-1} \hat{r}_{k,T}^{up} e^{ik\lambda}$$

and recalling that $\Upsilon_T(\lambda) = \sum_{k=-(T-1)}^{T-1} r_k e^{ik\lambda}$, we need to establish a concentration inequality on $\mathbb{P}\left[\sup_{\lambda \in [0,2\pi)} |\widehat{\Upsilon}_T^{up}(\lambda) - \Upsilon_T(\lambda)| > x\right]$. For any $\lambda \in [0, 2\pi)$, the term $\widehat{\Upsilon}_T^{up}(\lambda)$ can be written as (see Lemma 7)

$$
\begin{aligned}
\widehat{\Upsilon}_T^{up}(\lambda) &= d_T(\lambda)^{\mathsf{H}} \left(\frac{Y_T^{\mathsf{H}} Y_T}{N} \odot B_T \right) d_T(\lambda) \\
&= d_T(\lambda)^{\mathsf{H}} \left(\frac{V_T^{\mathsf{H}} V_T}{N} \odot B_T \right) d_T(\lambda) \\
&\quad + d_T(\lambda)^{\mathsf{H}} \left(\frac{P_T^{\mathsf{H}} V_T + V_T^{\mathsf{H}} P_T}{N} \odot B_T \right) d_T(\lambda) \\
&\quad + d_T(\lambda)^{\mathsf{H}} \left(\frac{P_T^{\mathsf{H}} P_T}{N} \odot B_T \right) d_T(\lambda) \\
&\triangleq \widehat{\Upsilon}_T^{u}(\lambda) + \widehat{\Upsilon}_T^{cross}(\lambda) + \widehat{\Upsilon}_T^{sig}(\lambda)
\end{aligned}
$$

where B_T is the matrix defined in the statement of Lemma 7. We know from the proof of Theorem 2 that

$$
\mathbb{P}\left[\sup_{\lambda \in [0,2\pi)} |\widehat{\Upsilon}_T^{u}(\lambda) - \Upsilon_T(\lambda)| > x \right] \le \exp\left(-\frac{cTx^2}{4 \|\boldsymbol{\Upsilon}\|_{\infty}^2 \log T}(1 + o(1)) \right). \tag{20}
$$

We then need only handle the terms $\widehat{\Upsilon}_T^{cross}(\lambda)$ and $\widehat{\Upsilon}_T^{sig}(\lambda)$.

We start with a simple lemma.

Lemma 13. *Let X and Y be two independent $\mathcal{N}(0,1)$ random variables. Then for any $\tau \in (-1, 1)$,*

$$
\mathbb{E}[\exp(\tau XY)] = (1 - \tau^2)^{-1/2}.
$$

Proof:

$$
\begin{aligned}
\mathbb{E}[\exp(\tau XY)] &= \frac{1}{2\pi} \int_{\mathbb{R}^2} e^{\tau xy} e^{-x^2/2} e^{-y^2/2} \, dx \, dy \\
&= \frac{1}{2\pi} \int_{\mathbb{R}^2} e^{-(x-\tau y)^2/2} e^{-(1-\tau^2)y^2/2} \, dx \, dy \\
&= (1 - \tau^2)^{-1/2}.
\end{aligned}
$$

$\blacksquare$

With this result, we now have

Lemma 14. *There exists a constant $a > 0$ such that*

$$
\mathbb{P}\left[\sup_{\lambda \in [0,2\pi)} |\widehat{\Upsilon}_T^{cross}(\lambda)| > x \right] \le \exp\left(-\frac{axT}{\sqrt{\log T}}(1 + o(1)) \right).
$$

Proof: We only sketch the proof of this lemma. We show that for any $\lambda \in [0, 2\pi]$,

$$
\mathbb{P}[|\widehat{\Upsilon}_T^{cross}(\lambda)| > x] \le \exp\left(-\frac{axT}{\sqrt{\log T}} + C \right)
$$

where C does not depend on $\lambda \in [0, 2\pi]$. The lemma is then proven by a discretization argument of the interval $[0, 2\pi]$ analogous to what was done in the proofs of Section II. We shall bound $\mathbb{P}[\widehat{\Upsilon}_T^{cross}(\lambda) > x]$, the term

$\mathbb{P}[\widehat{\Upsilon}_T^{cross}(\lambda) < -x]$ being bounded similarly. From Lemma 8, we get

$$\widehat{\Upsilon}_T^{cross}(\lambda) = \text{Tr}\Big(D_T(\lambda)^{\mathsf{H}} \frac{P_T^{\mathsf{H}} V_T + V_T^{\mathsf{H}} P_T}{N} D_T(\lambda) B_T \Big)$$

$$= \text{Tr}\, \frac{D_T(\lambda)^{\mathsf{H}} (\Gamma_T^{1/2})^{\mathsf{H}} \boldsymbol{s}_T \boldsymbol{h}_T^{\mathsf{H}} W_T R_T^{1/2} D_T(\lambda) B_T}{N}$$

$$+ \text{Tr}\, \frac{D_T(\lambda)^{\mathsf{H}} (R_T^{1/2})^{\mathsf{H}} W_T^{\mathsf{H}} \boldsymbol{h}_T \boldsymbol{s}_T^{\mathsf{H}} \Gamma_T^{1/2} D_T(\lambda) B_T}{N}$$

$$= \frac{2}{N} \Re(\boldsymbol{h}_T^{\mathsf{H}} W_T G_T(\lambda) \boldsymbol{s}_T)$$

where $G_T(\lambda) = R_T^{1/2} D_T(\lambda) B_T D_T(\lambda)^{\mathsf{H}} (\Gamma_T^{1/2})^{\mathsf{H}}$. Let $G_T(\lambda) = U_T \Omega_T \widetilde{U}_T^{\mathsf{H}}$ be a singular value decomposition of $G_T(\lambda)$ where $\Omega = \text{diag}(\omega_0, \dots, \omega_{T-1})$. Observe that the vector $\boldsymbol{x}_T \triangleq W_T^{\mathsf{H}} \boldsymbol{h}_T = (x_0, \dots, x_{T-1})^{\mathsf{T}}$ has the distribution $\mathcal{CN}(0, \|\boldsymbol{h}_T\|^2 I_T)$. We can then write

$$\widehat{\Upsilon}_T^{cross}(\lambda) \overset{\mathcal{L}}{=} \frac{2}{N} \Re\left(\boldsymbol{x}_T^{\mathsf{H}} \Omega_T \boldsymbol{s}_T \right) = \frac{2}{N} \sum_{t=0}^{T-1} \omega_t (\Re x_t \Re s_t + \Im x_t \Im s_t).$$

Notice that $\{\Re x_t, \Im x_t, \Re s_t, \Im s_t\}_{t=0}^{T-1}$ are independent with $\Re x_t, \Im x_t \sim \mathcal{N}(0, \|\boldsymbol{h}_T\|^2/2)$ and $\Re s_t, \Im s_t \sim \mathcal{N}(0, 1/2)$. Letting $0 < \tau < (\sup_T \|\boldsymbol{h}_T\|)^{-1} (\sup_\lambda \|G_T(\lambda)\|)^{-1}$ and using Markov's inequality and Lemma 13, we get

$$\mathbb{P}\left[\widehat{\Upsilon}_T^{cross}(\lambda) > x \right] = \mathbb{P}\left[e^{N\tau \widehat{\Upsilon}_T^{cross}(\lambda)} > e^{N\tau x} \right] \le e^{-N\tau x} \mathbb{E}\left[e^{2\tau \sum_t \omega_t (\Re x_t \Re s_t + \Im x_t \Im s_t)} \right]$$

$$= e^{-N\tau x} \prod_{t=0}^{T-1} \left(1 - \tau^2 \omega_t^2 \|\boldsymbol{h}_T\|^2 \right)^{-1} = \exp\left(-N\tau x - \sum_{t=0}^{T-1} \log(1 - \tau^2 \omega_t^2 \|\boldsymbol{h}_T\|^2) \right).$$

Mimicking the proof of Lemma 9, we can establish that $\sum_t \omega_t^2 = O(\log T)$ and $\max_t \omega_t = O(\sqrt{\log T})$ uniformly in $\lambda \in [0, 2\pi]$. Set $\tau = b/\sqrt{\log T}$ where $b > 0$ is small enough so that $\sup_{T,\lambda}(\tau \|\boldsymbol{h}_T\| \|G_T(\lambda)\|) < 1$. Observing that $\log(1 - x) = O(x)$ for x small enough, we get

$$\mathbb{P}[\widehat{\Upsilon}_T^{cross}(\lambda) > x] \le \exp\left(-Nbx/\sqrt{\log T} + \mathcal{E}(\lambda, T) \right)$$

where $|\mathcal{E}(\lambda, T)| \le (C/\log T) \sum_t \omega_t^2 \le C$. This establishes Lemma 14. ∎

Lemma 15. *There exists a constant $a > 0$ such that*

$$\mathbb{P}\left[\sup_{\lambda \in [0,2\pi)} |\widehat{\Upsilon}_T^{sig}(\lambda)| > x \right] \le \exp\left(-\frac{axT}{\sqrt{\log T}}(1 + o(1)) \right).$$

Proof: By Lemma 8,

$$\widehat{\Upsilon}_T^{sig}(\lambda) = N^{-1} \text{Tr}(D_T^{\mathsf{H}} P_T^{\mathsf{H}} P_T D_T B_T)$$

$$= \frac{\|\boldsymbol{h}_T\|^2}{N} \boldsymbol{s}_T^{\mathsf{H}} G_T(\lambda) \boldsymbol{s}_T$$

where $G_T(\lambda) = \Gamma_T^{1/2} D_T(\lambda) B_T D_T(\lambda)^{\mathsf{H}} (\Gamma_T^{1/2})^{\mathsf{H}}$. By the spectral factorization $G_T(\lambda) = U_T \Sigma_T U_T^{\mathsf{H}}$ with $\Sigma_T = \text{diag}(\sigma_0, \dots, \sigma_{T-1})$, we get

$$\widehat{\Upsilon}_T^{sig}(\lambda) \overset{\mathcal{L}}{=} \frac{\|\boldsymbol{h}_T\|^2}{N} \sum_{t=0}^{T-1} \sigma_t |s_t|^2$$

and

$$\mathbb{P}[\widehat{\Upsilon}_T^{sig}(\lambda) > x] \le e^{-N\tau x}\mathbb{E}\Big[e^{\tau\|\boldsymbol{h}_T\|^2 \sum_t \sigma_t |s_t|^2}\Big]$$

$$= \exp\Big(-N\tau x - \sum_{t=0}^{T-1} \log(1 - \sigma_t \tau \|\boldsymbol{h}_T\|^2)\Big)$$

for any $\tau \in (0, 1/(\|\boldsymbol{h}_T\|^2 \sup_\lambda \|G_T(\lambda)\|))$. Let us show that

$$|\operatorname{Tr} G_T(\lambda)| \le C\sqrt{\frac{\log T + 1}{T}}.$$

Indeed, we have

$$|\operatorname{Tr} G_T(\lambda)| = N^{-1}|\operatorname{Tr} D_T B_T D_T^{\mathsf{H}} \Gamma_T| = \frac{1}{N}\left|\sum_{k,\ell=0}^{T-1} \frac{e^{-\imath(k-\ell)\lambda}\gamma_{\ell,k}}{T - |k-\ell|}\right|$$

$$\le \Big(\frac{1}{N}\sum_{k,\ell=0}^{T-1} |\gamma_{k,\ell}|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{k,\ell=0}^{T-1} \frac{1}{(T - |k-\ell|)^2}\Big)^{1/2}$$

$$= \Big(\frac{\operatorname{Tr}\Gamma_T\Gamma_T^{\mathsf{H}}}{N}\Big)^{1/2}\Big(\frac{2}{N}(\log T + C)\Big)^{1/2} \le C\sqrt{\frac{\log T + 1}{T}}.$$

Moreover, similar to the proof of Lemma 9, we can show that $\sum_t \sigma_t^2 = O(\log T)$ and $\max_t |\sigma_t| = O(\sqrt{\log T})$ uniformly in λ. Taking $\tau = b/\sqrt{\log T}$ for $b > 0$ small enough, and recalling that $\log(1-x) = 1 - x + O(x^2)$ for x small enough, we get that

$$\mathbb{P}[\widehat{\Upsilon}_T^{sig}(\lambda) > x] \le \exp\Big(-\frac{Nbx}{\sqrt{\log T}} + \frac{b\|\boldsymbol{h}_T\|^2}{\sqrt{\log T}}\operatorname{Tr} G_T(\lambda) + \mathcal{E}(T,\lambda)\Big)$$

where $|\mathcal{E}(T,\lambda)| \le (C/\log T)\sum_t \sigma_t^2 \le C$. We therefore get

$$\mathbb{P}[\widehat{\Upsilon}_T^{sig}(\lambda) > x] \le \exp\Big(-\frac{Nbx}{\sqrt{\log T}} + C\Big)$$

where C is independent of λ. Lemma 15 is then obtained by the discretization argument of the interval $[0, 2\pi]$.

∎

Gathering Inequality (20) with Lemmas 14 and 15, we get the second inequality of the statement of Theorem 3.

IV. APPLICATION TO SOURCE DETECTION

Consider a sensor network composed of N sensors impinged by zero (hypothesis H_0) or one (hypothesis H_1) source signal. The stacked signal matrix $Y_T = [y_0, \dots, y_{T-1}] \in \mathbb{C}^{N \times T}$ from time $t = 0$ to $t = T - 1$ is modeled as

$$Y_T = \begin{cases} V_T & , H_0 \\ \boldsymbol{h}_T s_T^{\mathsf{H}} + V_T & , H_1 \end{cases} \tag{21}$$

where $s_T^{\mathsf{H}} = [s_0^*, \dots, s_{T-1}^*]$ are (hypothetical) independent $\mathcal{CN}(0,1)$ signals transmitted through the constant channel $\boldsymbol{h}_T \in \mathbb{C}^N$, and $V_T = W_T R_T^{1/2} \in \mathbb{C}^{N \times T}$ models a stationary noise matrix as in (1).

As opposed to standard procedures where preliminary pure noise data are available , we shall proceed here to an online signal detection test solely based on Y_T, by exploiting the consistence established in Theorem 3.

The approach consists precisely in estimating R_T by $\widehat{R}_T \in \{\widehat{R}_T^{bp}, \widehat{R}_T^{up}\}$, which is then used as a whitening matrix for Y_T. The binary hypothesis (21) can then be equivalently written

$$Y_T \widehat{R}_T^{-1/2} = \begin{cases} W_T R_T \widehat{R}_T^{-1/2} & , H_0 \\ \boldsymbol{h}_T \boldsymbol{s}_T^{\mathsf{H}} \widehat{R}_T^{-1/2} + W_T R_T \widehat{R}_T^{-1/2} & , H_1. \end{cases} \tag{22}$$

Since $\|R_T \widehat{R}_T^{-1} - I_T\| \to 0$ almost surely (by Theorem 3 as long as $\inf_{\lambda \in [0,2\pi)} \boldsymbol{\Upsilon}(\lambda) > 0$), for T large, the decision on the hypotheses (22) can be handled by the generalized likelihood ratio test (GLRT) [12] by approximating $W_T R_T \widehat{R}_T^{-1/2}$ as a purely white noise. We then have the following result.

Theorem 4. *Let $\widehat{R}_T$ be any of $\widehat{R}_T^{bp}$ or $\widehat{R}_T^{up}$ strictly defined in Theorem 3 for Y_T now following model* (21). *Further assume $\inf_{\lambda \in [0,2\pi)} \boldsymbol{\Upsilon}(\lambda) > 0$ and define the test*

$$\alpha = \frac{N \left\| Y_T \widehat{R}_T^{-1} Y_T^{\mathsf{H}} \right\|}{\mathrm{Tr}\left(Y_T \widehat{R}_T^{-1} Y_T^{\mathsf{H}} \right)} \underset{H_1}{\overset{H_0}{\lessgtr}} \gamma \tag{23}$$

where $\gamma \in \mathbb{R}^+$ satisfies $\gamma > (1 + \sqrt{c})^2$. Then, as $T \to \infty$,

$$\mathbb{P}\left[\alpha \geq \gamma\right] \to \begin{cases} 0 & , H_0 \\ 1 & , H_1. \end{cases}$$

Recall from [12] that the decision threshold $(1 + \sqrt{c})^2$ corresponds to the almost sure limiting largest eigenvalue of $\frac{1}{T} W_T W_T^{\mathsf{H}}$, that is the right-edge of the support of the Marčenko–Pastur law.

Simulations are performed hereafter to assess the performance of the test (23) under several system settings. We take here $\boldsymbol{h}_T$ to be the following steering vector $\boldsymbol{h}_T = \sqrt{p/T}[1, \ldots, e^{2i\pi\theta(T-1)}]$ with $\theta = 10°$ and p a power parameter. The matrix R_T models an autoregressive process of order 1 with parameter a, *i.e.* $[R_T]_{k,l} = a^{|k-l|}$.

In Figure 2, the detection error $1 - \mathbb{P}[\alpha \geq \gamma | H_1]$ of the test (23) for a false alarm rate (FAR) $\mathbb{P}[\alpha \geq \gamma | H_0] = 0.05$ under $\widehat{R}_T = \widehat{R}_T^{up}$ (Unbiased) or $\widehat{R}_T = \widehat{R}_T^{bp}$ (Biased) is compared against the estimator that assumes R_T perfectly known (Oracle), *i.e.* that sets $\widehat{R}_T = R_T$ in (23), and against the GLRT test that wrongly assumes temporally white noise (White), *i.e.* that sets $\widehat{R}_T = I_T$ in (23). The source signal power is set to $p = 1$, that is a signal-to-noise ratio (SNR) of 0 dB, N is varied from 10 to 50 and $T = N/c$ for $c = 0.5$ fixed. In the same setting as Figure 2, the number of sensors is now fixed to $N = 20$, $T = N/c = 40$ and the SNR (hence p) is varied from -10 dB to 4 dB. The powers of the various tests are displayed in Figure 3 and compared to the detection methods which estimate R_T from a pure noise sequence called Biased PN (pure noise) and Unbiased PN. The results of the proposed online method are close to that of Biase/Unbiased PN, this last presenting the disadvantage to have at its disposal a pure noise sequence at the receiver.

Both figures suggest a close match in performance between Oracle and Biased, while Unbiased shows weaker performance. The gap evidenced between Biased and Unbiased confirms the theoretical conclusions.
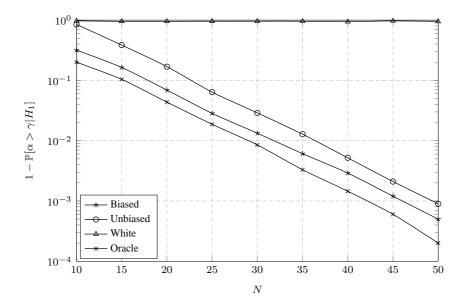
Figure 2. Detection error versus N with FAR= 0.05, $p = 1$, SNR= 0 dB, $c = 0.5$, and $a = 0.6$.
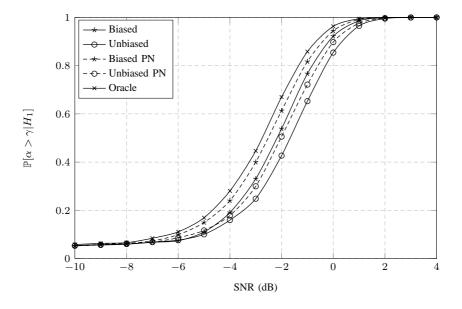

Figure 3. Power of detection tests versus SNR (dB) with FAR= 0.05, $N = 20$, $c = 0.5$, and $a = 0.6$.

APPENDIX

A. Proofs for Theorem 1

1) Proof of Lemma 3: Developing the quadratic forms given in the statement of the lemma, we get

$$d_T(\lambda)^{\mathsf{H}} \frac{V_T^{\mathsf{H}} V_T}{N} d_T(\lambda) = \frac{1}{NT} \sum_{l,l'=0}^{T-1} e^{-\imath(l'-l)\lambda} [V_T^{\mathsf{H}} V_T]_{l,l'}$$

$$= \frac{1}{NT} \sum_{l,l'=0}^{T-1} e^{-\imath(l'-l)\lambda} \sum_{n=0}^{N-1} v_{n,l}^* v_{n,l'}$$

$$= \sum_{k=-(T-1)}^{T-1} e^{-\imath k\lambda} \frac{1}{NT} \sum_{n=0}^{N-1} \sum_{t=0}^{T-1} v_{n,t}^* v_{n,t+k} \mathbb{1}_{0 \le t+k \le T-1}$$

$$= \sum_{k=-(T-1)}^{T-1} \hat{r}_k^b e^{-\imath k\lambda} = \widehat{\Upsilon}_T^b(\lambda),$$

and

$$\mathbb{E}\left[d_T(\lambda)^{\mathsf{H}} \frac{V_T^{\mathsf{H}} V_T}{N} d_T(\lambda) \right] = d_T(\lambda)^{\mathsf{H}} (R_T^{1/2})^{\mathsf{H}} \frac{\mathbb{E}[W_T^{\mathsf{H}} W_T]}{N} R_T^{1/2} d_T(\lambda)$$

$$= d_T(\lambda)^{\mathsf{H}} R_T d_T(\lambda).$$

B. Proofs for Theorem 2

1) Proof of Lemma 7: We have

$$d_T(\lambda)^{\mathsf{H}} \left(\frac{V_T^{\mathsf{H}} V_T}{N} \odot B_T \right) d_T(\lambda) = \frac{1}{NT} \sum_{l,l'=-(T-1)}^{T-1} e^{i(l-l')\lambda} [V_T^{\mathsf{H}} V_T]_{l,l'} \frac{T}{T - |l - l'|}$$

$$= \sum_{k=-(T-1)}^{T-1} e^{ik\lambda} \frac{1}{N(T-|k|)} \sum_{n=0}^{N-1} \sum_{t=0}^{T-1} v_{n,t}^* v_{n,t+k} \mathbb{1}_{0 \le t+k \le T-1}$$

$$= \sum_{k=-(T-1)}^{T-1} \hat{r}_k^u e^{ik\lambda} = \widehat{\Upsilon}_T^u(\lambda).$$

2) Proof of Lemma 9: We start by observing that

$$\mathrm{Tr}\, B_T^2 = \sum_{i,j=0}^{T-1} [B_T]_{i,j}^2 = \sum_{i,j=0}^{T-1} \left(\frac{T}{T - |i-j|} \right)^2 = 2 \sum_{i>j}^{T-1} \left(\frac{T}{T - |i-j|} \right)^2 + T$$

$$= 2 \sum_{k=1}^{T-1} \left(\frac{T}{T-k} \right)^2 (T-k) + T = 2T^2 \sum_{k=1}^{T-1} \frac{1}{T-k} + T = 2T^2 \left(\log T + C \right).$$

Inequality (10) is then obtained upon noticing that $\|B_T\| \le \sqrt{\mathrm{Tr}\, B_T^2}$.

We now show (11). Using twice the inequality $\mathrm{Tr}(FG) \le \|F\| \mathrm{Tr}(G)$ when $F, G \in \mathbb{C}^{m \times m}$ and G is nonnegative definite [10], we get

$$\sum_{t=0}^{T-1} \sigma_t^2(\lambda_i) = \mathrm{Tr}\, Q_T(\lambda_i)^2 = \mathrm{Tr}\, R_T D_T(\lambda_i) B_T D_T(\lambda_i)^{\mathsf{H}} R_T D_T(\lambda_i) B_T D_T(\lambda_i)^{\mathsf{H}}$$

$$\le \|R_T\| \, \mathrm{Tr}\, R_T (D_T(\lambda_i) B_T D_T(\lambda_i)^{\mathsf{H}})^2$$

$$\le T^{-2} \|R_T\|^2 \, \mathrm{Tr}(B_T^2) \le 2 \|\Upsilon\|_\infty^2 \log T + C.$$

Inequality (12) is immediate since $\|Q_T\|^2 \le \mathrm{Tr}\, Q_T^2$.

As regards (13), by the Cauchy–Schwarz inequality,

$$\sum_{t=0}^{T-1} |\sigma_t^3(\lambda_i)| = \sum_{t=0}^{T-1} \sigma_t^2(\lambda_i)|\sigma_t(\lambda_i)| \leq \sqrt{\sum_{t=0}^{T-1} \sigma_t^4(\lambda_i) \sum_{t=0}^{T-1} \sigma_t^2(\lambda_i)}$$

$$\leq \sqrt{\left(\sum_{t=0}^{T-1} \sigma_t^2(\lambda_i)\right)^2 \sum_{t=0}^{T-1} \sigma_t^2(\lambda_i)} = \left(\sum_{t=0}^{T-1} \sigma_t^2(\lambda_i)\right)^{3/2}$$

$$= C((\log T)^{3/2} + 1).$$

REFERENCES

[1] W. B. Wu and M. Pourahmadi, "Banding sample covariance matrices of stationary processes," *Statist. Sinica*, vol. 19, pp. 1755–1768, 2009.

[2] P. J. Bickel and E. Levina, "Regularized estimation of large covariance matrices," *Ann. Statist.*, vol. 36, pp. 199–227, 2008.

[3] C. Lam and J. Fan, "Sparsistency and rates of convergence in large covariance matrices estimation," *Ann. Statist.*, vol. 37, pp. 4254–4278, 2009.

[4] H. Xiao and W. B. Wu, "Covariance matrix estimation for stationary time series," *Ann. Statist.*, vol. 40, pp. 466–493, 2012.

[5] T. T. Cai, C.-H. Zhang, and H. H. Zhou, "Optimal rates of convergence for covariance matrix estimation," *Ann. Statist.*, vol. 38, pp. 2118–2144, 2010.

[6] T. T. Cai, Z. Ren, and H. H. Zhou, "Optimal rates of convergence for estimating Toeplitz covariance matrices," *Prob. Th. and Related Fields*, vol. 156, pp. 101–143, 2013.

[7] P. Vallet and P. Loubaton, "Toeplitz rectification and DoA estimation with MUSIC." Submitted to IEEE ICASSP'14, Florence, Italy, 2014.

[8] R. M. Gray, *Toeplitz and circulant matrices: A review*. Now Pub., 2006.

[9] O. Kallenberg, *Foundations of modern probability*, ser. Probability and its Applications (New York). New York: Springer-Verlag, 1997.

[10] R. A. Horn and C. R. Johnson, *Topics in matrix analysis*. Cambridge: Cambridge University Press, 1991.

[11] M. Ledoux, *The concentration of measure phenomenon*, ser. Mathematical Surveys and Monographs. Providence, RI: American Mathematical Society, 2001, vol. 89.

[12] P. Bianchi, M. Debbah, M. Maida, and J. Najim, "Performance of statistical tests for single-source detection using random matrix theory," *IEEE Transactions on Information Theory*, vol. 57, no. 4, pp. 2400–2419, 2011.